European Journal of Operational Research 237 (2014) 665–676

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

Coarse-grained optimization-driven design and piecewise linear modeling of synthetic genetic circuits

Ali R. Zomorrodi [a,b,*], Costas D. Maranas [a]

[a] Department of Chemical Engineering, The Pennsylvania State University, University Park, PA 16802, USA
[b] Department of Industrial & Manufacturing Engineering, The Pennsylvania State University, University Park, PA 16802, USA

ARTICLE INFO

Article history:
Received 5 January 2013
Accepted 25 January 2014
Available online 5 February 2014

Keywords:
(B) Bioinformatics
Synthetic genetic circuits
Piecewise linear differential equations
Integer linear program

ABSTRACT

The aim of synthetic biology is to confer novel functions to cells by rationally interconnecting basic genetic parts into circuits. A key barrier in the design of synthetic genetic circuits is that only a qualitative description of the performance and interactions of the basic genetic parts is available in databases such as the Registry of Standard Biological Parts. Modeling approaches capable of harnessing this qualitative knowledge are thus timely. Here, we introduce an optimization-based framework, which makes use of the available qualitative information about basic biological parts to automatically identify the circuit elements and structures enabling a desired response to the presence/absence of input signals. Promoters and ribosome binding sites are categorized as high, medium or low efficiency and protein expressions in the circuit are described using piecewise linear differential equations. The desired function of the circuit is also mathematically described as the maximization/minimization of a constrained objective function. We employed this framework for the design of a toggle switch, a genetic decoder and a genetic half adder unit. The identified designs are consistent with previously constructed circuit configurations and in some cases point to completely new architectures. The identified non-intuitive circuit structures highlight the importance of accounting for ribosome binding site efficiencies and relative protein abundance levels in circuit design. Our results reaffirm the usefulness of the qualitative information for the coarse-grained genetic circuit design and simulation in the absence of detailed quantitative information.

1. Introduction

Synthetic biology aims to combine engineering principles and molecular biology to design and construct biological systems with novel functions that cannot be found in nature. This is achieved by interconnecting basic biological parts into more complex modules and systems bringing about the desired functionality (i.e., synthetic genetic circuits) and subsequently transferring them into the cells. The pioneering efforts in this direction were the design and experimental construction of two fundamental synthetic genetic circuits namely the toggle switch (Gardner, Cantor, & Collins, 2000) and the repressilator (Elowitz & Leibler, 2000). The genetic toggle switch is a synthetic bistable gene-regulatory network (Gardner et al., 2000), whereas the repressilator is an oscillating network that periodically induces the synthesis of a fluorescent protein (Elowitz & Leibler, 2000). Several other researchers used these foundations to construct similar toggle switches in *Escherichia coli* (Atkinson, Savageau, Myers, & Ninfa, 2003) and mammalian cells (Kramer, Fischer, & Fussenegger, 2004), synchronized relaxation oscillators (McMillen, Kopell, Hasty, & Collins, 2002), a synthetic gene-metabolic oscillator (Fung et al., 2005), a synthetic genetic edge detection program (Tabor et al., 2009) and many others (Kim & Winfree, 2011; Levskaya et al., 2005; Moon, Lou, Tamsir, Stanton, & Voigt, 2012; Purcell, Savery, Grierson, & di Bernardo, 2010; Stanton et al., 2013). There have been also other efforts for constructing synthetic cell–cell communication circuits (Basu, Gerchman, Collins, Arnold, & Weiss, 2005; Brenner, Karig, Weiss, & Arnold, 2007; Choi, Ha, Park, & Kim, 2011; Song, Payne, Tan, & You, 2011).

Synthetic genetic circuits were shown to have potential applications in many different areas ranging from metabolic engineering and biofuel production to drug discovery, medicine and bio-sensing (Anderson, Clarke, Arkin, & Voigt, 2006; Chen & Weiss, 2005; Dellomonaco, Fava, & Gonzalez, 2010; Medema, Breitling, Bovenberg, & Takano, 2011; Neumann & Neumann-Staubitz, 2010; Shen & Liao, 2011; Xie, Wroblewska, Prochazka, Weiss, & Benenson, 2011). For example, Chen and Weiss (2005) designed a synthetic cell–cell communication system in yeast that can eliminate the

* Corresponding author at: Department of Chemical Engineering, The Pennsylvania State University, University Park, PA 16802, USA. Tel.: +1 814 863 9958.
E-mail address: alirzomorrodi@gmail.com (A.R. Zomorrodi).

http://dx.doi.org/10.1016/j.ejor.2014.01.054
0377-2217/

need for monitoring of the fermentation processes and the control of gene expression through addition of expensive inducers (Andrianantoandro, Basu, Karig, & Weiss, 2006). Applications of synthetic biology in metabolic engineering is concerned with designing novel and synthetic genetic and metabolic circuits redirecting carbon flow toward the product of interest such as biofuels. For example, Shen and Liao (2011) constructed a synthetic iterative pathway in *E. coli* for 2-ketoacid elongation. In a more recent study Zhang, Carothers, and Keasling (2012) developed a dynamic sensor-regulator system for biodiesel production in *E. coli*. Potential applications of synthetic biology in medicine were also demonstrated by engineering the interaction between bacteria and mammalian cells to enable *E. coli* to invade specific cancer-derived cells (Anderson et al., 2006). In another study, Xie et al. (2011) engineered a synthetic genetic circuit capable of triggering the cell death upon the detection of the cancerous gene expression patterns while leaving the non-cancerous cells unharmed. The ever-increasing applications of synthetic genetic circuits spurred the construction of the Registry of Standard Biological Parts (http://partsregistry.org/) in an effort toward the compilation and standardization of the building blocks of these circuits. This registry is a collection of various basic genetic components (i.e., biobricks) such as promoters (regulatory regions), protein coding regions and ribosome binding sites (RBSs), as well as composite parts such as logic gates. The development of this database has enabled the design and construction of new integrated genetic circuits through the assembly of standardized interchangeable biological parts in analogy to the use of parts in electrical circuits.

In line with these rapid developments in the standardization, experimental construction and applications of synthetic genetic circuits, development of new computational and modeling procedures to assist the design process and verify the behavior of the genetic circuits prior to their experimental implementation is an absolute necessity. A number of previous studies aimed to address this need (Chandran, Bergmann, & Sauro, 2009; Chandran, Bergmann, & Sauro, 2010; Chandran & Sauro, 2012; Marchisio & Stelling, 2008, 2011; Weeding, Houle, & Kaznessis, 2010). For example, Marchisio and Stelling (2008) proposed a framework to simulate the behavior of synthetic genetic circuits with 'composable' parts using ordinary differential equations, where different parts can be inter-connected through exchange of common signal carriers such as polymerases per second (PoPS) and ribosomes per second (RiPS). In a subsequent study, the same researchers combined this approach with the so-called Karnaugh map for the design of electrical circuits to automate the design of digital biological circuits (Marchisio & Stelling, 2011). Weeding et al. (2010) also introduced a software tool called SynBioSS designer for the quantitative simulation of biological circuits, which takes as input the list of components in a given circuit (e.g., biobricks from the Registry of Standard Biological Parts) and automatically generates a kinetic model of biological interactions involved. In addition, following the successful application of operations research and optimization-based techniques in medicine and biology, and in particular, for the analysis of natural gene regulatory networks (Cordone & Lulli, 2013; Dasika, Gupta, & Maranas, 2004, 2005; Gerhard-Wilhelm, Kropat, Akteke-Öztürk, & Görgülü, 2009; Kergosien, Tournamille, Laurence, & Billaut, 2011; Liu, Chen, & Chen, 2013; Mazier, Billaut, & Tournamille, 2010; Palafox, Noman, & Iba, 2013; Yang, Maguire, Yarmush, & Androulakis, 2008), a few efforts have been made toward using these techniques for the optimal design of synthetic genetic circuits. For example, Dasika and Maranas (2008) introduced OptCircuit, a mixed-integer nonlinear dynamic optimization framework that relies on the availability of kinetic parameters describing the function and interaction of circuit components to identify the required circuit elements and the structure that fulfill a pre-specified task. A customized outer-approximation method was used in this study to bracket the optimal solution of this dynamic optimization problem (Dasika & Maranas, 2008). In a more recent study Huynh, Kececioglu, Koppe, and Tagkopoulos (2012) used a linear approximation of the nonlinear terms around a desired steady-state to solve the mixed-integer dynamic optimization problem of the circuit design more efficiently.

Despite these efforts, a key roadblock in the comprehensive utilization of the basic parts available in databases such as the Registry of Standard Biological Parts and elsewhere is that, in contrast to electronic circuits, biological components largely lack quantitative information characterizing their function and interactions. Instead, only a qualitative description of basic parameters of the biobricks and the interactions among them is usually available. This is due to not only the imprecise understanding of their function but more fundamentally due to the context-specific nature of their action. For example, only a qualitative description (low/medium/high) of promoter strength or efficiency of RBS and only descriptive information about the interaction between regulatory proteins and promoters, or ligands and regulatory proteins is available in this registry. Given this qualitative description of the basic part characteristics, a key question is how to systematically assemble these components into complex circuits with desired responses to various external stimuli. Computational approaches to tackle this question are still nonexistent. To fill this gap, here we introduce an optimization-based framework combined with a piecewise-linear differential equation (De Jong et al., 2004; Glass, 1975; Glass & Kauffman, 1973) description of protein expression for identifying the circuit elements and structure(s) to meet a user-specified functionality. We applied this framework to a variety of applications including the design of a toggle switch, a 2–4 bit genetic decoder (a circuit that detects the presence/absence of two external biological signals) and a 1-bit genetic half adder unit. Some of the identified circuits are consistent with previously constructed circuits, whereas some are novel non-intuitive structures.

2. Biological background

2.1. Gene expression

DNA is a large biological molecule with a double helix structure carrying inherited information, whereas genes are functional segments of the DNA coding for proteins. Proteins are large chains of amino acids considered as the most fundamental functional units in a cell performing a wide range of biological functions such as catalyzing metabolic reactions, DNA replication, gene expression regulation, transferring molecules to and from the cell and many more. Gene expression is the process of synthesizing a functional protein by decoding the genetic information in the respective gene and is composed of two fundamental processes, namely, transcription and translation. Transcription uses a specific gene on DNA as a template to produce a single-stranded molecule, called messenger RNA (mRNA). This is followed by translation, which involves the conversion of genetic instructions encoded by mRNA into amino acids and chaining amino acids together in a specific order (using a biological machine called ribosome) to synthesize the protein (Gene → mRNA → protein). Ribosome binding site (RBS) is a sequence on mRNA that binds to ribosome when initiating protein translation.

2.2. Control of the gene expression

Cells respond to any environmental stimulus, such as the presence of a specific compound in the growth medium through regulating the gene expression events. Gene expression regulation can occur at transcription, translation or post-translation levels